[8] ——, "Geometrical and numerical design of structured unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3722–3735, Aug. 2006.
[9] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.
[10] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*. New York: Academic [A subsidiary of Harcourt Brace Jovanovich], 1978, vol. 80, Pure and Applied Mathematics.
[11] O. Henkel, "Sphere packing bounds in the Grassmann and Stiefel manifolds," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3445–3456, Oct. 2005.
[12] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.
[13] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.
[14] X.-B. Liang and X.-G. Xia, "Unitary signal constellations for differential space-time modulation with two transmit antennas: Parametric codes, optimal designs and bounds," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2291–2322, Aug. 2002.
[15] L. Santaló, "Integral Geometry and Geometric Probability," in *Encyclopedia of Mathematics and its Applications*. Reading, MA: Addison-Wesley, 1976, vol. 1, Section: Probability.
[16] A. Shokrollahi, B. Hassibi, B. M. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2335–2367, Sep. 2001.

Subchannel Allocation in Multiuser Multiple-Input–Multiple-Output Systems

Pedro Tejera, Wolfgang Utschick, *Senior Member, IEEE*, Gerhard Bauch, *Senior Member, IEEE*, and Josef A. Nossek, *Fellow, IEEE*

***Abstract*—Assuming perfect channel state information at the transmitter of a Gaussian broadcast channel, strategies are investigated on how to assign subchannels in frequency and space domain to each receiver aiming at a maximization of the sum rate transmitted over the channel. For the general sum capacity maximizing solution, which has recently been found, a method is proposed that transforms each of the resulting vector channels into a set of scalar channels. This makes possible to achieve capacity by simply using scalar coding and detection techniques. The high complexity involved in the computation of this optimum solution motivates the introduction of a novel suboptimum zero-forcing allocation strategy that directly results in a set of virtually decoupled scalar channels. Simulation results show that this technique tightly approaches the performance of the optimum solution, i.e., complexity reduction comes at almost no cost in terms of sum capacity. As the optimum solution, the zero-forcing allocation strategy applies to any number of transmit antennas, receive antennas and users.**

***Index Terms*—Broadcast channel, multiuser multiple-input multiple-output (MIMO), orthogonal frequency division multiplexing (OFDM), successive encoding, sum capacity, zero-forcing.**

I. Introduction

Increasing demand for broadband services calls for higher data rates in future wireless communication systems [1]. Data rates of several Mb/s for high mobility scenarios and up to 1 Gb/s in low mobility or static scenarios are expected in fourth generation systems. In the way to such transmission rates there are two major barriers to be overcome. The first is the scarcity of spectrum, which limits the amount of bandwidth available for transmission. The second is the wireless channel that severely distorts the signal due to multipath propagation.

The combination of multiple antennas and multicarrier technology seems to be ideal to achieve the expected rates under the mentioned constraints [2]. On the one hand, multiple-input multiple-output (MIMO) channels resulting from the use of multiple antennas at both transmitter and receiver show higher capacity than single-input–single-output (SISO) channels and this difference linearly grows for increasing transmit power. Thus, multiple antennas lead to higher spectral efficiency. On the other hand, multicarrier techniques, such as orthogonal frequency division multiplexing (OFDM), transform the frequency selective broadband channel into a set of nearly flat narrowband channels. As a result, distortion due to multipath is reduced and equalization at the receiver is greatly simplified.

In the work at hand, we consider the downlink of a wireless communication system with multiple antennas at the transmitter and the receivers and OFDM as transmission scheme. We assume that receivers

Manuscript received November 29, 2004; revised May 3, 2006. The material in this correspondence was presented in part at the IEEE International Conference on Communications, Istanbul, Turkey, June 2006.

P. Tejera, W. Utschick, and J. A. Nossek are with the Institute for Circuit Theory and Signal Processing, Munich University of Technology, 80290 Munich, Germany (e-mail: tejera@nws.ei.tum.de; utschick@nws.ei.tum.de; nossek@nws.ei.tum.de@nws.ei.tum.de).

G. Bauch is with the DoCoMo Communications Laboratories Europe GmbH, 80687 Munich, Germany (e-mail: bauch@docomolab-euro.com).

Communicated by R. R. Müller, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2006.881751

know their respective transmission channels perfectly and the transmitter has perfect knowledge of the channel of every user. This assumption presumes perfect channel estimates, a quasi-static scenario, in which channels do not change significantly for the time between two consecutive channel estimations, and channel reciprocity in case of a time division duplex (TDD) system or a feedback link in case of a frequency division duplex (FDD) system.[1] Under these assumptions we investigate how to allocate resources among the users, viz. power, space and frequency, in order to maximize the sum of rates transmitted to the receivers in the cell.

The downlink scenario with perfect channel-state information at the transmitter (CSIT) has been extensively investigated in recent years from different points of view. Assuming users with single receive antennas, linear and non linear precoding solutions have been proposed for criteria such as minimum transmit power subject to quality of service constraints imposed by the mobile terminals [4], [5] or sum of mean square errors (MSE) at the receivers with limited transmit power and with or without zero-forcing constraints [6], [7]. An overview on precoding approaches with non cooperative receive antennas can be found in [8]. Users with multiple antennas have been considered in [9] and [10], where beamforming algorithms have been proposed aiming at a minimization of the transmit power under quality constraints for transmission on each link. Following the same or related criteria, a multiuser MIMO OFDM setting has been discussed in [11], [12] and [13]. As pointed out in [11], a major open issue in such a setting is subchannel allocation, i.e., over how many links should each particular user receive information and at which frequencies.

From the viewpoint of information theory the downlink of a wireless communication system is a broadcast channel. For our purposes, since multiple antennas are available at the transmitter the channel is non degraded and since noise at the receivers is Gaussian the channel is also Gaussian [14]. Contrary to the degraded broadcast channel, for the nondegraded broadcast channel the capacity region has remained unknown until very recently. The possibility to employ multiple antennas at the base station has motivated the resumption of work related to nondegraded broadcast channels in the last years. This work is largely based on results obtained in the late seventies and early eighties. In [15], it was shown that, assuming single receive antennas and for the two-user case, a successive encoding technique based on the coding technique with known interference presented in [16] reaches the Sato upperbound [17] on sum capacity for the broadcast channel. The achievability region obtained by successively encoding users, which happens to be a particular instance of Marton's region [18], was extended in [19] to the case of multiple users and multiple antennas. This region has been shown to reach the Sato bound and the transmit covariance matrices achieving the points on its boundary have been characterized based on duality results between the broadcast and multiple access channels [20]. Recently, it has been found that the successive encoding region is in fact the actual capacity region of the Gaussian broadcast channel [21]. Computation of the sum capacity achieving covariance matrices requires solving a convex optimization problem in the multiple access channel and transforming the solution back to the dual broadcast channel. For the computation of the covariance matrices that achieve sum capacity, algorithms have been found in [22]–[24] that solve the first step iteratively. Also motivated by these information theoretic results some work has recently been done considering criteria other than sum capacity, e.g., [25]–[27].

In the present work, practical aspects of the optimum solution for the sum capacity of a Gaussian nondegraded broadcast channel given in [20] are discussed. In particular, we transform the resulting vector channels into a set of scalar subchannels while preserving capacity and discuss the extension of this solution to OFDM. Note that conversion of vector channels to scalar subchannels is always beneficial as it simplifies encoding and decoding of information. Due to the relatively high complexity associated with the optimum solution for increasing numbers of subcarriers and users, we develop a zero-forcing subchannel allocation method that simplifies computation of transmit covariance matrices as well as downlink signaling while keeping close to the Sato bound. The method proceeds sequentially assigning at each step a spatial subchannel to a certain user so that no interference is caused on already established subchannels. Interference caused by a certain subchannel on subsequently established subchannels is neutralized by successive encoding. This is possible by choosing the encoding order to be the same as the order in which subchannels are assigned. The result is a set of virtually decoupled subchannels over which capacity can be maximized by distributing transmit power according to the waterfilling solution.

If all receive antennas are able to cooperate, our zero-forcing allocation method is equivalent to a singular value decomposition (SVD) of the MIMO channel, which does not require the successive encoding feature. On the contrary, if no cooperation among receive antennas is possible, our method is equivalent to a zero-forcing with successive encoding approach (ZF-SE) (e.g., [28], [14], [29]) with optimized encoding order and user selection. In the intermediate cases, in which only groups of antennas can cooperate with each other, the MIMO channels corresponding to groups of cooperative receive antennas are effectively diagonalized. Due to the ability to exploit the cooperation capability of receive antennas and the successive allocation of subchannels to users, we call our approach cooperative zero-forcing with successive encoding and successive allocation method (CZF-SESAM).

With OFDM as transmission scheme, the algorithm can be applied to every subcarrier simultaneously. If the number of transmit antennas is exceeded by the number of users or receive antennas, the algorithm jointly selects spatial dimensions, encoding order and users to be served at each frequency, thus performing a kind of optimized subcarrier allocation as a by-product.

The remainder of the correspondence is organized as follows. Section II describes the system model employed in this work. In Section III, the optimum solution for the sum capacity of a Gaussian nondegraded broadcast channel is reviewed and practical aspects of this solution are discussed. In Section IV, we start reviewing the ZF-SE and SVD approaches, which apply to the case of non cooperative and cooperative receive antennas, respectively. Combining the principles of these two approaches and elaborating on the encoding order issue we arrive at our CZF-SESAM approach and discuss its main properties. In Section V, simulation results are shown and commented and, finally, the correspondence is concluded and the main results are summarized in Section VI.

Notation

In the following, vectors and matrices are denoted by lower case bold and capital bold letters, respectively. Random variables are represented with sans-serif characters. We use $(\bullet)^*$ for complex conjugation, $(\bullet)^{\mathrm{T}}$ for matrix transposition and $(\bullet)^{\mathrm{H}}$ for conjugate transposition. $\mathrm{E}\{\bullet\}$ and $\mathrm{Tr}\{\bullet\}$ denote the expectation and trace operators, respectively. Given a matrix $\boldsymbol{A}$, $[\boldsymbol{A}]_{i,\bullet}$ represents its ith row, $[\boldsymbol{A}]_{\bullet,i}$ its ith column and $|\boldsymbol{A}|$ its determinant. For Hermitian matrices, $\boldsymbol{A} \geq 0$ indicates that matrix $\boldsymbol{A}$ is positive semidefinite. Letting $\{\boldsymbol{A}_i\}_{i=1,\ldots,I}$ be the set of all matrices indexed by the variable i, $\mathrm{diag}[\boldsymbol{A}_1,\ldots,\boldsymbol{A}_I]$ represents a

[1]The reader is referred to [3] for further references on the subject of channel knowledge acquisition at the transmitter.

block diagonal matrix with matrices $\{\boldsymbol{A}_i\}_{i=1,\ldots,I}$ as blocks in the main diagonal. Finally, the identity matrix of dimension q is denoted by $\boldsymbol{I}_q$.

II. System Model

A. Channel Model

We consider the downlink of a cellular wireless communication system. The base station is equipped with t transmit antennas. Each user $k \in \{1,\ldots,K\}$ has r_k receive antennas. An OFDM transmission scheme is employed with a cyclic prefix that is assumed to be longer than the length of the power delay profile of the channel so that no intersymbol interference (ISI) occurs. The channel is assumed to be invariant for the duration of an OFDM symbol so that orthogonality between subcarriers is preserved during transmission and no intercarrier interference (ICI) occurs. According to these assumptions the relationship between the vector of transmitted signals $\boldsymbol{x}(n) \in \mathbb{C}^{t\times 1}$ and the vector $\boldsymbol{y}_k(n) \in \mathbb{C}^{r_k\times 1}$ of receive signals for user k at subcarrier $n \in \{1,\ldots,N\}$ can be expressed as

$$\boldsymbol{y}_k(n) = \boldsymbol{H}_k(n)\boldsymbol{x}(n) + \boldsymbol{w}_k(n) \tag{1}$$

where $\boldsymbol{H}_k(n) \in \mathbb{C}^{r_k\times t}$ is the channel matrix seen by user k at subcarrier n and $\boldsymbol{w}_k(n) \in \mathbb{C}^{r_k\times 1}$ a realization of a zero-mean circularly symmetric complex Gaussian distributed random variable $\mathbf{w}_k(n)$ representing noise with covariance matrix $\mathrm{E}\{\mathbf{w}_k(n)\mathbf{w}_k(n)^{\mathrm{H}}\} = \boldsymbol{I}_{r_k}$. Noise processes of different subcarriers are considered uncorrelated. The transmitter is assumed to perfectly know all matrices $\boldsymbol{H}_k(n)$ and the average transmit power over the whole spectrum is limited, i.e.

$$\frac{1}{N}\sum_{n=1}^{N}\mathrm{Tr}\left\{\mathrm{E}\{\mathbf{x}(n)\mathbf{x}(n)^{\mathrm{H}}\}\right\} \le P_{\mathrm{Tx}}. \tag{2}$$

For notational convenience, the subcarrier index n will be omitted whenever it is not essential for the understanding of the discussion. For any subcarrier, stacking receive and noise vectors and the channel matrices corresponding to every user, i.e., $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathrm{T}} \ \ldots \ \boldsymbol{y}_K^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{w} = [\boldsymbol{w}_1^{\mathrm{T}} \ \ldots \ \boldsymbol{w}_K^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{H} = [\boldsymbol{H}_1^{\mathrm{T}} \ \ldots \ \boldsymbol{H}_K^{\mathrm{T}}]^{\mathrm{T}}$, the following expression is obtained that indicates the relationship between all signals in the downlink at that frequency:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}. \tag{3}$$

In the following, we call $\boldsymbol{H}$ the composite channel matrix to distinguish it from the individual channel matrices of each of the users. As the conditional probability density function $p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{y}|\boldsymbol{x})$ is Gaussian, model (3) represents a Gaussian broadcast channel [30], over which we assume that independent information is transmitted to each user, i.e., no common information is broadcast to all users.

B. Structure of the Transmit Signal

As far as the capacity limits of the nondegraded Gaussian broadcast channel are concerned, the appropriate dependence of the actually transmitted signals, viz. $\boldsymbol{x}$, on the signals intended for each user, as well as the optimum way to generate these have for long time been unclear. In [15] a linear relationship between transmit and intended signals was proposed. This dependence together with a successive encoding of information was proved to be optimum in terms of sum capacity for $K = 2$ and single receive antennas, i.e., $r_1 = r_2 = 1$. Posterior work [20], [21] has shown that this way of generating and structuring the transmit signal is optimum in terms of capacity for any number of receive antennas and users.

Based on these results the transmit signal can be written as

$$\boldsymbol{x} = \sum_{k=1}^{K}\boldsymbol{V}_k\boldsymbol{P}_k^{1/2}\boldsymbol{s}_k \tag{4}$$

where $\boldsymbol{V}_k \in \mathbb{C}^{t\times m_k}$ is a matrix with orthonormal column vectors, $\boldsymbol{P}_k \in \mathbb{R}^{m_k\times m_k}$ is a diagonal power matrix and $\boldsymbol{s}_k \in \mathbb{C}^{m_k\times 1}$ is the vector of signals intended for user k, which is assumed to be a realization of a zero-mean, circularly symmetric complex Gaussian distributed vector $\mathbf{s}_k$ with covariance $\mathrm{E}\{\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}\} = \boldsymbol{I}_{m_k}$ and statistically independent of signals intended for other users. The number of spatial dimensions m_k is less than or equal to $\min\{t, r_k\}$ and in order to satisfy [2]

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\mathrm{Tr}\{\boldsymbol{P}_k(n)\} \le P_{\mathrm{Tx}}. \tag{5}$$

At each frequency, signals $\boldsymbol{s}_k$ are obtained as result of a successive encoding of information intended for the different users. Let $\pi_n : \{1,\ldots,K\} \to \{1,\ldots,K\}$ be a bijective function assigning to each user the order in which its intended information sequence is encoded on subcarrier n. According to (1) and (4) the signal received by user $\pi_n(1)$ is given by

$$\boldsymbol{y}_{\pi_n(1)} = \boldsymbol{H}_{\pi_n(1)}\boldsymbol{V}_{\pi_n(1)}\boldsymbol{P}_{\pi_n(1)}^{1/2}\boldsymbol{s}_{\pi_n(1)} + \underbrace{\boldsymbol{H}_{\pi_n(1)}\sum_{i>1}\boldsymbol{V}_{\pi_n(i)}\boldsymbol{P}_{\pi_n(i)}^{1/2}\boldsymbol{s}_{\pi_n(i)} + \boldsymbol{w}_{\pi_n(1)}}_{\boldsymbol{z}_{\pi_n(1)}}.$$

Since at the time of encoding information for user $\pi_n(1)$, signals $\boldsymbol{s}_{\pi_n(i>1)}$ are unknown, they can not be taken into account for the generation of $\boldsymbol{s}_{\pi_n(1)}$ and will add to the Gaussian noise to yield a total noise term $\boldsymbol{z}_{\pi_n(1)}$. Correspondingly, the mutual information between the receive and transmit signals for user $\pi_n(1)$ can be written as

$$I(\mathbf{y}_{\pi_n(1)};\mathbf{s}_{\pi_n(1)}) = \log\frac{\left|\boldsymbol{R}_{\pi_n(1)} + \boldsymbol{H}_{\pi_n(1)}\boldsymbol{Q}_{\pi_n(1)}\boldsymbol{H}_{\pi_n(1)}^{\mathrm{H}}\right|}{\left|\boldsymbol{R}_{\pi_n(1)}\right|} \tag{6}$$

where

$$\boldsymbol{Q}_k = \boldsymbol{V}_k\boldsymbol{P}_k\boldsymbol{V}_k^{\mathrm{H}} \tag{7}$$

and

$$\boldsymbol{R}_{\pi_n(j)} = \boldsymbol{I}_{r_{\pi_n(j)}} + \boldsymbol{H}_{\pi_n(j)}\sum_{i>j}\boldsymbol{Q}_{\pi_n(i)}\boldsymbol{H}_{\pi_n(j)}^{\mathrm{H}}. \tag{8}$$

User $\pi_n(j)$, for which information is encoded in the jth place, receives a signal vector

$$\begin{aligned}\boldsymbol{y}_{\pi_n(j)} = {} & \boldsymbol{H}_{\pi_n(j)}\boldsymbol{V}_{\pi_n(j)}\boldsymbol{P}_{\pi_n(j)}^{1/2}\boldsymbol{s}_{\pi_n(j)} \\ & + \boldsymbol{H}_{\pi_n(j)}\sum_{i<j}\boldsymbol{V}_{\pi_n(i)}\boldsymbol{P}_{\pi_n(i)}^{1/2}\boldsymbol{s}_{\pi_n(i)} \\ & + \underbrace{\boldsymbol{H}_{\pi_n(j)}\sum_{i>j}\boldsymbol{V}_{\pi_n(i)}\boldsymbol{P}_{\pi_n(i)}^{1/2}\boldsymbol{s}_{\pi_n(i)} + \boldsymbol{w}_{\pi_n(j)}}_{\boldsymbol{z}_{\pi_n(j)}}.\end{aligned} \tag{9}$$

Now, at the time of encoding information for user $\pi_n(j)$, signals $\boldsymbol{s}_{\pi_n(i<j)}$ are known and can be taken into account in the encoding process to generate the signal $\boldsymbol{s}_{\pi_n(j)}$. Theoretically, such a way of encoding information allows to achieve a transmission rate as high as the capacity of a channel in which the second term of (9) were not present

[16], [19]. Accordingly, the mutual information between receive and intended signals for user $\pi_n(j)$ reads

$$I(\mathbf{y}_{\pi_n(j)}; \mathbf{s}_{\pi_n(j)}) = \log \frac{\left|\boldsymbol{R}_{\pi_n(j)} + \boldsymbol{H}_{\pi_n(j)}\boldsymbol{Q}_{\pi_n(j)}\boldsymbol{H}^{\mathrm{H}}_{\pi_n(j)}\right|}{\left|\boldsymbol{R}_{\pi_n(j)}\right|} \tag{10}$$

Note that for the last encoded user $\boldsymbol{R}_{\pi_n(K)} = \boldsymbol{I}_{r_{\pi_n(K)}}$, i.e., the mutual information is completely determined by its own covariance matrix and the noise process at the receiver and it is not affected by signals intended for other users. We also observe that, if $\mathrm{E}\left\{\mathbf{w}_k\mathbf{w}_k^{\mathrm{H}}\right\} = \boldsymbol{W}_k$, noise whitening at the receiver does not affect mutual information, i.e., noise whitening preserves capacity. Therefore, there is no loss of generality in assuming $\mathrm{E}\{\mathbf{w}_k\mathbf{w}_k^{\mathrm{H}}\} = \boldsymbol{I}_{r_k}$, as it was done at the beginning of this section.

III. Sum Capacity of the Broadcast Channel

A. Known Solution

For a frequency flat channel, single carrier transmission, i.e., $N = 1$, and a particular choice of encoding order and transmit covariance matrices the sum of achievable rates is given by

$$\begin{aligned} &C(\pi, \boldsymbol{H}, \boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K) \\ &\quad= \sum_{k=1}^{K} \log \frac{\left|\boldsymbol{I}_{r_{\pi(k)}} + \boldsymbol{H}_{\pi(k)} \sum_{i\geq k} \boldsymbol{Q}_{\pi(i)}\boldsymbol{H}^{\mathrm{H}}_{\pi(k)}\right|}{\left|\boldsymbol{I}_{r_{\pi(k)}} + \boldsymbol{H}_{\pi(k)} \sum_{i> k} \boldsymbol{Q}_{\pi(i)}\boldsymbol{H}^{\mathrm{H}}_{\pi(k)}\right|} \end{aligned} \tag{11}$$

where π is the function that defines the order in which users are encoded.

Maximization of (11) is possible over the choice of encoding order and covariance matrices subject to the transmit power constraint (5). However, since (11) is neither a convex nor a concave function of the covariance matrices, direct optimization will generally involve an exhaustive search over the entire space of covariance matrices that satisfy the power constraint and over the set of encoding orders. An alternative method to solve this problem has been found in [20] that exploits the close relationship between the capacity region of the broadcast channel and that of its dual multiple access channel.

Given a broadcast channel as described by (3), the system model for the dual multiple access channel reads

$$\boldsymbol{t} = \sum_{k=1}^{K} \boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{r}_k + \boldsymbol{w},$$

where $\boldsymbol{t} \in \mathbb{C}^{t\times 1}$ is the vector of receive signals, $\boldsymbol{r}_k \in \mathbb{C}^{r_k\times 1}$ is the vector of signals transmitted by user k, which is assumed to be a realization of a zero-mean circularly symmetric complex Gaussian distributed random variable $\mathbf{r}_k$ with covariance matrix $\mathrm{E}\{\mathbf{r}_k\mathbf{r}_k^{\mathrm{H}}\} = \boldsymbol{\Sigma}_k$, and $\boldsymbol{w} \in \mathbb{C}^{t\times 1}$ is a realization of a zero-mean circularly symmetric complex Gaussian distributed noise process with unit covariance matrix. Under a collective constraint on the sum of transmit powers, i.e.

$$\sum_{k=1}^{K} \mathrm{Tr}\{\boldsymbol{\Sigma}_k\} \leq P_{\mathrm{Tx}}, \tag{12}$$

it has been shown in [20] that the set of rates achievable in the multiple access channel by successively decoding users, which is optimum in terms of capacity, is equal to the set of rates achievable in the dual broadcast channel by performing a successive encoding of users. Moreover, given a set of covariance matrices and a particular decoding order, a method has been found to compute the covariance matrices that achieve the same rates in the broadcast channel by encoding users in reverse order, i.e., the user decoded first in the multiple access channel is encoded last in the broadcast channel. Note that the multiple access channel with constraint (12) is merely a mathematical tool that allows computation of optimum operational points in the broadcast channel. Obviously, a common power constraint shared by non-cooperating users lacks practical relevance.

As a consequence of this result a maximization of (11) can be indirectly performed by first maximizing the sum of achievable rates in the dual multiple access channel and then computing the covariance matrices that achieve that sum rate in the broadcast channel. Fortunately, the sum of achievable rates in the multiple access channel, given by

$$C(\boldsymbol{H}, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K) = \log\left|\boldsymbol{I}_t + \sum_{k=1}^{K}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{\Sigma}_k\boldsymbol{H}_k\right|, \tag{13}$$

is a concave function of the covariance matrices and, therefore, can be maximized by using convex optimization techniques.

From the optimum set of covariance matrices in the multiple access channel, which does not depend on decoding order, $K!$ different sets of covariance matrices for the broadcast channel can be obtained that maximize the sum of achievable rates. Each of these sets corresponds to a different encoding order. In [20] it has been shown that the maximum value of (11) subject to (12) reaches the Sato upperbound on sum capacity for a Gaussian broadcast channel, i.e., successive encoding achieves the sum capacity in this kind of channels.

B. Practical Issues

Successive encoding relies on the possibility to completely neutralize interference that is only known at the transmitter by using adequate codes. The existence of such codes was first proved in [16] for scalar channels and this result has recently been extended to vector channels by several authors (see [19] and references therein). However, these proofs use random codes that lack algebraic structure and detectors based on statistical typicality, which is difficult to implement.

More convenient in practical terms are suboptimum coding techniques that are able to counteract known interference with reduced complexity such as the trellis codes presented in [19], [31] or the general lattice codes dicussed in [32] and [33] of which the popular Tomlinson-Harashima technique [34], [35] is a particular case. A common characteristic of these techniques is that they are exclusively applicable to scalar channels. As a consequence, it is covenient to look for a method that, for a given set of covariance matrices and encoding order, converts the resulting vector channels into sets of orthogonal scalar channels while preserving the rate of each of the users. In this way, capacity can be approached by using existing coding techniques for scalar channels.

In the following, we propose a method that decomposes the vector channel of each user into a set of orthogonal scalar channels while preserving capacity. Recalling the channel model for the jth encoded user (cf. (9)) we first apply a linear zero-forcing filter $\boldsymbol{F}_{\pi(j)}$ at the receiver. The output signal reads

$$\begin{aligned}\boldsymbol{y}'_{\pi(j)} = \boldsymbol{s}_{\pi(j)} + \boldsymbol{F}_{\pi(j)}\boldsymbol{H}_{\pi(j)}\sum_{i<j}\boldsymbol{B}_{\pi(i)}\boldsymbol{s}_{\pi(i)} \\ \underbrace{+\boldsymbol{F}_{\pi(j)}\left(\boldsymbol{H}_{\pi(j)}\sum_{i>j}\boldsymbol{B}_{\pi(i)}\boldsymbol{s}_{\pi(i)} + \boldsymbol{w}_{\pi(j)}\right)}_{\boldsymbol{z}_{\pi(j)}}\end{aligned}$$

where for notational covenience $\boldsymbol{B}_{\pi(i)} = \boldsymbol{V}_{\pi(i)}\boldsymbol{P}^{1/2}_{\pi(i)}$ and the zero-forcing filter $\boldsymbol{F}_{\pi(j)}$ is given by

$$\boldsymbol{F}_{\pi(j)} = \left(\boldsymbol{B}^{\mathrm{H}}_{\pi(j)}\boldsymbol{H}^{\mathrm{H}}_{\pi(j)}\boldsymbol{R}^{-1}_{\pi(j)}\boldsymbol{H}_{\pi(j)}\boldsymbol{B}_{\pi(j)}\right)^{-1}\boldsymbol{B}^{\mathrm{H}}_{\pi(j)}\boldsymbol{H}^{\mathrm{H}}_{\pi(j)}\boldsymbol{R}^{-1}_{\pi(j)}.$$

Note that this linear transformation of the receive signal preserves the rate of user $\pi(j)$ and, as it only applies to the receiver of that user, it does not affect the rates of any other user.

The covariance matrix of the effective noise at the output of the zero-forcing filter $z'_{\pi(j)} = \boldsymbol{F}_{\pi(j)} z_{\pi(j)}$ equals

$$\boldsymbol{R}'_{\pi(j)} = \left(\boldsymbol{B}^{\mathrm{H}}_{\pi(j)} \boldsymbol{H}^{\mathrm{H}}_{\pi(j)} \boldsymbol{R}^{-1}_{\pi(j)} \boldsymbol{H}_{\pi(j)} \boldsymbol{B}_{\pi(j)}\right)^{-1}. \tag{14}$$

Performing an eigenvalue decomposition of this matrix, $\boldsymbol{R}'_{\pi(j)} = \boldsymbol{U}_{\pi(j)} \boldsymbol{\Lambda}_{\pi(j)} \boldsymbol{U}^{\mathrm{H}}_{\pi(j)}$, matrix $\boldsymbol{U}^{\mathrm{H}}_{\pi(j)}$ can be applied at the receiver to decorrelate the effective noise and signals can be transmitted along the column vectors of matrix $\boldsymbol{U}_{\pi(j)}$, i.e., $\boldsymbol{s}_{\pi(j)} = \boldsymbol{U}_{\pi(j)} \boldsymbol{s}'_{\pi(j)}$. As a result, the equivalent channel

$$\boldsymbol{y}''_{\pi(j)} = \boldsymbol{s}'_{\pi(j)} + \underbrace{\left(\boldsymbol{U}^{\mathrm{H}}_{\pi(j)} \boldsymbol{z}'_{\pi(j)}\right)}_{\boldsymbol{z}''_{\pi(j)}} + \boldsymbol{U}^{\mathrm{H}}_{\pi(j)} \boldsymbol{F}_{\pi(j)} \boldsymbol{H}_{\pi(j)} \sum_{i<j} \boldsymbol{B}_{\pi(i)} \boldsymbol{s}_{\pi(i)}$$

is obtained where the effective noise $z''_{\pi(j)}$ is uncorrelated and whose capacity can be achieved by separately coding over each of the scalar components. Note that correlation of the third term does not matter as this term is known at coding time and, therefore, its effect on performance can be completely nullified. As the transformation applied to decorrelate the effective noise is invertible the rate achievable by user $\pi(j)$ is preserved and, as it only applies to the receiver of that user, it does not affect the rates of any other user. As the statistics of $\mathbf{s}_{\pi(j)}$ are invariant under any unitary transformation neither the rate of user $\pi(j)$ nor the rate of any other user is affected by this kind of precoding.

The second issue discussed in this section refers to the extension of the sum capacity solution presented in [20] to an OFDM transmission scheme, i.e., $N > 1$. This extension is straightforward if for each user we arrange all frequency and space components in a unique block diagonal matrix

$$\tilde{\boldsymbol{H}}_k = \mathrm{diag}[\boldsymbol{H}_k(1), \ldots, \boldsymbol{H}_k(N)] \in \mathbb{C}^{Nr_k \times Nt}$$

and consider the channel model

$$\begin{bmatrix} \tilde{\boldsymbol{y}}_1 \\ \tilde{\boldsymbol{y}}_2 \\ \vdots \\ \tilde{\boldsymbol{y}}_K \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{\boldsymbol{H}}_1 \\ \tilde{\boldsymbol{H}}_2 \\ \vdots \\ \tilde{\boldsymbol{H}}_K \end{bmatrix}}_{\tilde{\boldsymbol{H}}} \tilde{\boldsymbol{x}} + \begin{bmatrix} \tilde{\boldsymbol{w}}_1 \\ \tilde{\boldsymbol{w}}_2 \\ \vdots \\ \tilde{\boldsymbol{w}}_K \end{bmatrix}$$

where $\tilde{\boldsymbol{w}}_k = [\boldsymbol{w}^{\mathrm{T}}_k(1) \quad \boldsymbol{w}^{\mathrm{T}}_k(2) \quad \ldots \quad \boldsymbol{w}^{\mathrm{T}}_k(N)]^{\mathrm{T}} \in \mathbb{C}^{Nr_k \times 1}$ and $\tilde{\boldsymbol{y}}_k \in \mathbb{C}^{Nr_k \times 1}$ and $\tilde{\boldsymbol{x}} \in \mathbb{C}^{Nt \times 1}$ are defined analogously. With such a model of the multicarrier broadcast channel the sum capacity maximizing transmit covariance matrices can be found by directly applying the method described above. Specifically, the sum of achievable rates in the corresponding dual multiple access channel is given by

$$C(\tilde{\boldsymbol{H}}, \tilde{\boldsymbol{\Sigma}}_1, \ldots, \tilde{\boldsymbol{\Sigma}}_K) = \log \left| \boldsymbol{I}_{Nt} + \sum_{k=1}^{K} \tilde{\boldsymbol{H}}^{\mathrm{H}}_k \tilde{\boldsymbol{\Sigma}}_k \tilde{\boldsymbol{H}}_k \right| \tag{15}$$

where $\tilde{\boldsymbol{\Sigma}}_k \in \mathbb{C}^{Nr_k \times Nr_k}$ is the transmit covariance matrix of user k. Using the same reasoning as in the proof of the Hadarmard's inequality given in [36] it is straightforward to show that matrices $\tilde{\boldsymbol{\Sigma}}_k$ optimally have a block diagonal structure matching the structure of their respective channels $\tilde{\boldsymbol{H}}_k$, i.e.

$$\tilde{\boldsymbol{\Sigma}}_k = \mathrm{diag}\left[\boldsymbol{\Sigma}_k(1), \boldsymbol{\Sigma}_k(2), \ldots, \boldsymbol{\Sigma}_k(N)\right] \in \mathbb{C}^{Nr_k \times Nr_k}$$

where $\boldsymbol{\Sigma}_k(n) \in \mathbb{C}^{r_k \times r_k}$. As a consequence, the sum of achievable rates given in (15) can optimally be expressed as sum of sum rates achieved on each subcarrier of the system

$$C(\tilde{\boldsymbol{H}}, \tilde{\boldsymbol{\Sigma}}_1, \ldots, \tilde{\boldsymbol{\Sigma}}_K) = \sum_{n=1}^{N} \log \left| \boldsymbol{I}_t + \sum_{k=1}^{K} \boldsymbol{H}^{\mathrm{H}}_k(n) \boldsymbol{\Sigma}_k(n) \boldsymbol{H}_k(n) \right|. \tag{16}$$

Since this expression is an addition of concave functions of matrices $\boldsymbol{\Sigma}_k(n)$, it is itself concave and can be maximized subject to the power constraint

$$\frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \mathrm{Tr}\{\boldsymbol{\Sigma}_k(n)\} \leq P_{\mathrm{Tx}}$$

by using convex optimization techniques. From the optimum transmit covariance matrices that apply to a particular subcarrier in the multiple access channel, a set of transmit covariance matrices that achieve the same sum rate on that subcarrier with equal transmit power in the dual broadcast channel can be computed by assuming a certain decoding order and using the transformations presented in [20]. Since there are $K!$ different decoding orders, there will be $K!$ such sets. The same transformations can be performed on every subcarrier assuming different decoding orders, which yields a set of KN transmit covariance matrices that achieve the sum capacity of the multicarrier broadcast channel. Note that, although there are $(K!)^N$ different optimum sets of transmit covariance matrices, there is a unique power allocation over frequency that achieves the sum capacity.

C. Complexity

Maximization of (13) or (16) under their respective power constraints falls into the category of determinant maximization problems with linear matrix inequality constraints. This kind of problems is at the top of a hierarchy of standard convex optimization problems including, among others, semidefinite programming or linear programming as especial cases [37]. In order to avoid the complexity of the very general numerical techniques that solve this kind of problems, iterative algorithms have been proposed in [22], [23] and [24] that guarantee convergence to the optimum solution. In the following we consider the iterative waterfilling approach presented in [22] as a representative of this kind of algorithms, which show similar properties and performance.[2] An extension of this algorithm to multicarrier transmission is sketched in Table I. In the first step, computation of effective channels, which result from noise whitening at the receiver, requires the inversion of KN matrices. The optimization problem in the second step can be solved by diagonalizing the effective channels and waterfilling the available transmit power over the resulting eigenvalues. This calls for KN eigenvalue decompositions and a waterfilling computation over $N\sum_{k=1}^{K} q_k$ dimensions with $q_k = \min\{r_k, t\}$. These two steps must be iteratively repeated until convergence is reached. Subsequently, a conversion of the solution to the dual broadcast channel must be performed that involves $2NK$ matrix inversions, as many matrix factorizations and NK SVDs. Additionally, if the method to obtain scalar channels described above is applied, further NK matrix inversions and as many eigenvalue decompositions are required. The memory required by these iterative approaches in order to store provisional results is in the best case $O(NK)$ (cf. [24]).

Finally, considering signaling, we observe that, in order to allow optimum detection, the transmitter should communicate to each receiver

[2] A detailed comparison of these algorithms is found in [23] and [24].

TABLE I
ITERATIVE ALGORITHM FOR MAXIMIZATION OF SUM CAPACITY

`initialization:` $\ell = 1, \quad \boldsymbol{\Sigma}_k^0(n) = \mathbf{0}, \ \forall k, n$
`repeat:`
1. $\boldsymbol{H}_k^{\text{eff},\ell}(n) = \boldsymbol{H}_k(n)\left(\boldsymbol{I}_t + \sum_{\substack{k'=1\\k'\neq k}}^{K} \boldsymbol{H}_k^{\text{H}}(n)\boldsymbol{\Sigma}_k^{\ell-1}(n)\boldsymbol{H}_k(n)\right)^{-1/2}, \ \forall n, k$
2. $\{\boldsymbol{M}_k^\ell(n)\}_{\substack{n=1,\cdots,N\\k=1,\cdots,K}} = \underset{\{\boldsymbol{A}_k^\ell(n)\}_{\substack{n=1,\cdots,N\\k=1,\cdots,K}}}{\arg\max} \sum_{n=1}^N \sum_{k=1}^K \log\left\lvert \boldsymbol{I}_t + \left(\boldsymbol{H}_k^{\text{eff},\ell}(n)\right)^{\text{H}} \boldsymbol{A}_k^\ell(n)\boldsymbol{H}_k^{\text{eff},\ell}(n)\right\rvert$
subject to $\frac{1}{N}\sum_{n=1}^N \sum_{k=1}^K \text{Tr}\left\{\boldsymbol{A}_k^\ell(n)\right\} \leq P_{\text{Tx}}, \ \boldsymbol{A}_k^\ell(n) \geq 0, \ \forall k, n$
3. $\boldsymbol{\Sigma}_k^\ell(n) = \left((K-1)\boldsymbol{\Sigma}_k^{\ell-1}(n) + \boldsymbol{M}_k^\ell(n)\right)/K, \ \forall k, n, \quad \ell = \ell + 1$
`until` convergence

k at least its corresponding transmit covariance matrices $\boldsymbol{Q}_k(n)$ and covariance matrices of effective noise $\boldsymbol{R}_k(n)$ on every subcarrier.

IV. Cooperative Zero-Forcing With Successive Encoding and Successive Allocation

In this section a technique is presented that needs a maximum of KNt SVDs to find an encoding order and a set of unit-norm transmit and receive weighting vectors that transform the broadcast channel into a set of virtually decoupled scalar channels over which the achievable sum rate is nearly equal to the sum capacity of the original channel. Complexity is similar to that of one iteration in any of the iterative approaches mentioned in the last section. Thus, as far as computational complexity is concerned, the basic saving consists in the fact that no iterations are required. Moreover, the solution directly applies to the broadcast channel and can be readily implemented using coding and detection techniques for scalar channels. In contrast to iterative approaches, the algorithm only needs $O(N)$ memory positions for storage of provisional results. Finally, in order to perform optimum detection, the transmitter must communicate to each user only its corresponding covariance matrix, which leads to a reduction of signaling overhead in the downlink [38]. This is a consequence of the zero-forcing constraint that the algorithm employs to compute transmit weighting vectors.

In the following, we first review and discuss two standard techniques on which our method is based. These are the zero-forcing with successive encoding technique (ZF-SE), which applies to channels with non-cooperating single receive antennas, and the SVD, which applies to single user channels in which antenna elements at the receiver can cooperate. A direct combination of both techniques results in a technique that we call block zero-forcing with successive encoding (block-ZF-SE). This technique eliminates inter-user interference following a ZF-SE strategy and exploits cooperation between antennas belonging to a same user. As all these techniques can be trivially applied to OFDM by parallel execution on every subcarrier, in the remaining of this section we assume a single carrier system, i.e., $N = 1$.

A. ZF-SE and SVD

Consider the system model given in (3) and assume that users have single receive antennas and do not cooperate with each other. A standard transmission technique that applies to this setting is ZF-SE (e.g., [28], [14], [29]).

Considering a structure of the transmit signal as indicated by (4) and an encoding order defined by a function π, the signal received by user $\pi(j)$ reads

$$\begin{aligned} y_{\pi(j)} = &\boldsymbol{h}_{\pi(j)}^{\text{T}}\boldsymbol{v}_{\pi(j)}p_{\pi(j)}^{1/2}s_{\pi(j)} \\ &+ \boldsymbol{h}_{\pi(j)}^{\text{T}}\sum_{i<j}\boldsymbol{v}_{\pi(i)}p_{\pi(i)}^{1/2}s_{\pi_n(i)} \\ &+ \underbrace{\boldsymbol{h}_{\pi(j)}^{\text{T}}\sum_{i>j}\boldsymbol{v}_{\pi(i)}p_{\pi(i)}^{1/2}s_{\pi(i)} + w_{\pi(j)}}_{z_{\pi(j)}} \end{aligned}$$

where $\boldsymbol{h}_k^{\text{T}} = [\boldsymbol{H}]_{k,\bullet}$. The ZF-SE technique selects vectors $\boldsymbol{v}_{\pi(i>j)}$ so that $z_{\pi(j)} = w_{\pi(j)}$, i.e., it completely suppresses interference caused by subsequently encoded users. The remaining intereference caused by previously encoded users is neutralized by coding. Computation of vectors $\boldsymbol{v}_k$ can be done either by permuting the rows of matrix $\boldsymbol{H}$ and performing an LQ factorization on the resulting matrix or in a successive way according to a Gram-Schmidt procedure as follows:

$$\boldsymbol{v}_{\pi(j)} = \frac{\boldsymbol{T}_j\boldsymbol{h}_{\pi(j)}^*}{\|\boldsymbol{T}_j\boldsymbol{h}_{\pi(j)}^*\|}, \quad j = 1, \ldots, q$$

where $\boldsymbol{T}_1 = \boldsymbol{I}_t, \boldsymbol{T}_{j+1} = \boldsymbol{T}_j - \boldsymbol{v}_{\pi(j)}\boldsymbol{v}_{\pi(j)}^{\text{H}}$ and q is the rank of $\boldsymbol{H}$. It is easy to prove that $\boldsymbol{T}_j$ is the projector matrix corresponding to the subspace defined by the intersection of kernels of channels $\boldsymbol{h}_{\pi(i<j)}^{\text{T}}$ and, as a consequence, $\boldsymbol{h}_{\pi(i<j)}^{\text{T}}\boldsymbol{v}_{\pi(j)} = 0$. On the other hand, in general, $\boldsymbol{h}_{\pi(i>j)}^{\text{T}}\boldsymbol{v}_{\pi(j)} \neq 0$, which constitutes the part of interference that can be combatted by coding. Effective transmission of information occurs over the set of virtually decoupled channels $\{\boldsymbol{h}_{\pi(j)}^{\text{T}}\boldsymbol{v}_{\pi(j)}\}_{j=1,\ldots,q}$ and the achievable sum rate is given by

$$C(\pi, \boldsymbol{H}, p_1, \ldots, p_q) = \sum_{j=1}^{q}\log\left(1 + p_{\pi(j)}g_{\pi(j)}^2\right) \tag{17}$$

where $g_{\pi(j)} = \left|\boldsymbol{h}_{\pi(j)}^{\text{T}}\boldsymbol{v}_{\pi(j)}\right|$. Subject to a transmit power constraint, (17) becomes maximum if power allocation is performed according to the waterfilling strategy.

Now, we may consider again model (3) and assume that K receivers with single antennas are able to cooperate or, equivalently, there is only a single user in the system with K receive antenna elements. In this case, it is well known that capacity is maximized by transmitting information along the right singular vectors of the channel matrix and allocating power according to a waterfilling strategy over the associated singular values [39]. At the receiver, though not essential, the receive signal can be filtered with the left singular vectors of the channel, which greatly simplifies detection. Capacity for this cooperative setting is given by

$$C(\boldsymbol{H}, p_1, \ldots, p_q) = \sum_{k=1}^{q} \log\left(1 + p_k \lambda_k^2\right) \tag{18}$$

where $\{\lambda_k\}_{k=1,\ldots,q}$ are the singular values of matrix $\boldsymbol{H}$.

It seems intuitively obvious that (18) give higher capacity values than (17). Indeed, if cooperation between receive antennas were detrimental, the receiver would always have the option to process receive signals independently as in (17). In [14] it has been shown that for $q = K$ both capacity values converge at high SNR values, i.e., under these conditions ZF-SE is optimum at high SNR values. This makes of ZF-SE an attractive option if the complexity of the optimum solution described above is to be reduced. However, ZF-SE as found in the literature only applies to single non-cooperating receive antennas.

Finally, we consider model (3) in all its generality, i.e., $r_k \geq 1, \forall\, k$. This is an intermediate case of the two discussed above. Antenna elements at the same receiver can cooperate, whereas between antenna elements belonging to different users cooperation is not possible. For such a setting we could do without the cooperation capability of groups of receive antennas and directly apply the ZF-SE approach. Nevertheless, improved performance is obtained if cooperation is exploited by combining the ZF-SE principle with the SVD. While the first combats inteference between groups of antennas belonging to different users, the second converts the vector channels of single users into a set of orthogonal scalar channels while preserving capacity. The essential difference between both a direct application of ZF-SE and a combination of ZF-SE and SVD, which we call block-ZF-SE, is that while ZF-SE eliminates the cross-talk between antennas of a same user, block-ZF-SE exploits this cross-talk in order to increase capacity.

Transmit and receive weighting vectors for the block-ZF-SE approach can be sequentially computed as follows:

$$\boldsymbol{H}_{\pi(j)}\boldsymbol{T}_j = \boldsymbol{U}_{\pi(j)}\boldsymbol{\Lambda}_{\pi(j)}\boldsymbol{V}_{\pi(j)}^{\mathrm{H}}, \quad j = 1, \ldots, \rho \tag{19}$$

where the right hand side of (19) is the SVD of the matrix on the left hand side, $\boldsymbol{V}_{\pi(j)} \in \mathbb{C}^{t \times m_{\pi(j)}}$ and $\boldsymbol{U}_{\pi(j)}^{\mathrm{H}} \in \mathbb{C}^{m_{\pi(j)} \times r_{\pi(j)}}$ are the matrices of transmit weighting unit-norm column vectors and receive weighting unit-norm row vectors of user $\pi(j)$, respectively and $m_{\pi(j)}$ is the rank of matrix $\boldsymbol{H}_{\pi(j)}\boldsymbol{T}_j$, where $\boldsymbol{T}_1 = \boldsymbol{I}_t$ and $\boldsymbol{T}_{j+1} = \boldsymbol{T}_j - \boldsymbol{V}_{\pi(j)}\boldsymbol{V}_{\pi(j)}^{\mathrm{H}}$. Finally, ρ is chosen such that $\sum_{j=1}^{\rho} m_{\pi(j)} = q$. Analogous to the ZF-SE approach, $\boldsymbol{T}_j$ is the projector matrix corresponding to the subspace defined by the intersection of kernels of the matrices $\boldsymbol{H}_{\pi(i<j)}$. Accordingly, considering the signal received by user $\pi(j)$ after weighting

$$\begin{aligned}\boldsymbol{y}_{\pi(j)} =\, & \boldsymbol{U}_{\pi(j)}^{\mathrm{H}}\boldsymbol{H}_{\pi(j)}\boldsymbol{V}_{\pi(j)}\boldsymbol{P}_{\pi(j)}^{1/2}\boldsymbol{s}_{\pi(j)} \\ & + \boldsymbol{U}_{\pi(j)}^{\mathrm{H}}\boldsymbol{H}_{\pi(j)}\sum_{i<j}\boldsymbol{V}_{\pi(i)}\boldsymbol{P}_{\pi(i)}^{1/2}\boldsymbol{s}_{\pi(i)} \\ & + \underbrace{\boldsymbol{U}_{\pi(j)}^{\mathrm{H}}\left(\boldsymbol{H}_{\pi(j)}\sum_{i>j}\boldsymbol{V}_{\pi_n(i)}\boldsymbol{P}_{\pi_n(i)}^{1/2}\boldsymbol{s}_{\pi_n(i)} + \boldsymbol{w}_{\pi_n(j)}\right)}_{\boldsymbol{z}_{\pi(j)}}\end{aligned}$$

we observe that $\boldsymbol{z}_{\pi(j)} = \boldsymbol{U}_{\pi(j)}^{\mathrm{H}}\boldsymbol{w}_{\pi_n(j)}$, i.e., interference caused by users $\pi(i > j)$ is linearly suppressed by the choice of transmit weighting vectors. The remaining interference, caused by users $\pi(i < j)$, can be rendered ineffective by coding. The effective channel for user $\pi(j)$ is a set of orthogonal scalar channels given by the diagonal matrix $\boldsymbol{\Lambda}_{\pi(j)}$, which has the same capacity as the projected channel $\boldsymbol{H}_{\pi(j)}\boldsymbol{T}_j$. Note that applying the ZF-SE approach to this projected channel would necessarily lead to a certain capacity loss as pointed out before when comparing (17) and (18). It can be easily verified that our block-ZF-SE approach converges to a SVD of the channel matrix if $K = 1$ and to a ZF-SE approach if $r_1 = \cdots = r_K = 1$.

Aiming at a maximization of sum rate there are two parameters of the block-ZF-SE approach, over which optimization can be done. As described above, at each step, the algorithm assigns to a certain user as many subchannels as the rank of its projected channel matrix. This is clearly suboptimum if, for instance, some of the subchannels are weak. In that case, contribution of these subchannels to the sum rate might be negligible while they may impose severe constraints on subchannels of subsequently encoded users. The second parameter is encoding order. In the following, coming back to the ZF-SE approach, an algorithm is proposed that orders users aiming at a maximization of the achievable sum rate. Since the algorithm involves a successive allocation of subchannels to users we call this approach zero-forcing with successive encoding and successive allocation method (ZF-SESAM). The criterion employed for the allocation of subchannels in this algorithm together with the allocation of just a single subchannel to a particular user at each step of the block-ZF-SE approach form the basis of our CZF-SESAM approach discussed later in this section.

B. ZF-SE and Encoding Order

We consider again model (3) with $r_1 = \cdots = r_K = 1$ and a ZF-SE transmission approach. As already mentioned, the achievable sum rate given in (17) is maximized by distributing power over the resulting subchannels according to a waterfilling strategy

$$p_{\pi(j)} = \max\left\{\eta - \frac{1}{g_{\pi(j)}^2}, 0\right\} \tag{20}$$

where $g_{\pi(j)} = |\boldsymbol{h}_{\pi(j)}^{\mathrm{T}}\boldsymbol{v}_{\pi(j)}|$ is the channel gain of user $\pi(j)$ and η is the waterfilling level, which is chosen to fulfil the transmit power constraint $p_{\pi(1)} + \cdots + p_{\pi(K)} \leq P_{\mathrm{Tx}}$. Substituting (20) into (17) under consideration of the transmit power constraint, the following expression results for the sum capacity of the system

$$\begin{aligned}C(\pi, \boldsymbol{H}) = Q\log\left(\frac{1}{Q}\left(P_{\mathrm{Tx}} + \sum_{\substack{j=1\\ p_{\pi(j)}\neq 0}}^{K}\frac{1}{g_{\pi(j)}^2}\right)\right) \\ + \log\left(\prod_{\substack{j=1\\ p_{\pi(j)}\neq 0}}^{K} g_{\pi(j)}^2\right)\end{aligned} \tag{21}$$

where $Q \leq K$ indicates the number of users for which $p_{\pi(j)} \neq 0$, i.e., the number of users that are served by the base station. As we observe, in contrast to the optimum solution, sum rate turns out to be a function of the encoding order for the ZF-SE approach.

Encoding order has been studied in [29] for $Q = K$, i.e., full rank matrix $\boldsymbol{H}$, $K \leq t$ and enough power so that every user can be served. Under this assumption it can be shown that the second term in (21) is independent of encoding order and sum rate is maximized by solving

$$\underset{\pi}{\operatorname{argmax}} \sum_{j=1}^{K}\frac{1}{g_{\pi(j)}^2}. \tag{22}$$

Based on this insight, in [29] a successive ordering algorithm is proposed that is shown to be optimum for $K = 2$. First, the last encoded user is chosen to be the one with minimum channel gain under orthogonality constraints imposed by all other users. Then, the user encoded before the last one is chosen to be the one with minimum channel gain

TABLE II
SUCCESSIVE ALLOCATION ALGORITHM FOR ZF-SESAM

```
initialization :   j = 1,   T_1 = I_t
repeat :
    1.  g_k = |h_k^T T_j|   ∀k
    2.  k_0 = argmax_k {g_k},   π(j) = k_0
    3.  T_{j+1} = T_j − T_j h*_{π(j)} h^T_{π(j)} T_j^H / g^2_{π(j)},   j = j + 1
until   j > q
```

under ortgonality constraints imposed by all other users except for the last one. The algorithm continues analogously until all users are ordered. This algorithm operates similarly to the V-BLAST [40] algorithm for detection of spatially multiplexed signals but with a minimum gain criterion instead of a maximum SNR criterion for the selection of a user at each step. While this ordering method can straightforwardly be applied for $K \leq t$, the case $K > t$ is not addressed in [29].

Here, we propose a subchannel allocation algorithm for the ZF-SE approach that aims at a maximization of the achievable sum rate and readily applies to any number of transmit antennas and users. This algorithm is sketched in Table II. In the first loop, the first encoded user is chosen to be the one with largest channel gain. In the second loop, considering the nullspace of the channel of the first encoded user, the second encoded user is chosen to be the one that exhibits the largest gain in this subspace. At any step of the algorithm, the user is selected that exhibits the largest gain within the subspace orthogonal to the channels of previously selected users. Due to the successive allocation of spatial channels to users we call this algorithm zero-forcing with successive encoding and successive allocation method (ZF-SESAM). The following two theorems provide some rationale for this ordering.

Theorem 1: Let $\mathcal{S}_j \subset \{1,\ldots,K\}$ be the set of first j selected users and let $C_{\mathcal{S}_j}$ be the sum rate achieved by these users. Selecting the next user according to the proposed method yields the maximum capacity increment $\Delta C = C_{\mathcal{S}_{j+1}} - C_{\mathcal{S}_j}$, where $\mathcal{S}_{j+1} = \mathcal{S}_j \cup \{\pi(j+1)\}$.

Proof: See Appendix. □

While the first theorem tells which user must be chosen to get the largest capacity increment, the second gives the optimum order for the selection of an additional pair of two arbitrary users.

Theorem 2: Let $\mathcal{S}_j \subset \{1,\ldots,K\}$ be the set of first j selected users and $k_1, k_2 \in \{1,\ldots,K\} \setminus \mathcal{S}_j$ so that $\|\boldsymbol{h}_{k_1}^{\mathrm{T}}\boldsymbol{T}_{j+1}\| \geq \|\boldsymbol{h}_{k_2}^{\mathrm{T}}\boldsymbol{T}_{j+1}\|$. Define $\mathcal{S}_{j+2} = \mathcal{S}_j \cup \{k_1, k_2\}$ and the ordering functions π and π' so that $\pi(i) = \pi'(i)\ \forall\, i \leq k$, $\pi(j+1) = \pi'(j+2) = k_1$ and $\pi(j+2) = \pi'(j+1) = k_2$. $C_{\mathcal{S}_{j+2}}$ is maximized by choosing the encoding order defined by π.

Proof: See Appendix. □

While these two theorems do provide a rationale for our ZF-SESAM algorithm, they do not prove its optimality. In fact, the final order resulting from the application of the algorithm might not be optimum. At each step the set of ordered users is optimally incremented provided that the order of previously selected users in the set is not altered. However, reconsidering this order could yield a better solution.

For $K \leq t$, if $K = 2$, it can be shown that the ordering obtained by applying this method is equal to that obtained in [29]. By contrast, if $K > 2$ orderings will in general be different. Simulation results with $t = 4$ show that, in this case, our method exhibits better performance in the low SNR region whereas the ordering proposed in [29] outperforms the method presented here in the high SNR region. However, in both regions, performance difference is in the order of some hundredths of a bit.

C. CZF-SESAM

The basic idea behind the successive allocation in our ZF-SESAM approach is the selection at each step of the spatial subchannel with largest gain within the subspace orthogonal to previously established subchannels. This idea can straightforwardly be applied to the general model (3) with $r_k \geq 1$ in combination with our block-ZF-SE algorithm. The resulting approach is called CZF-SESAM.

Contrary to the block-ZF-SE, the CZF-SESAM approach proceeds assigning at each step only one scalar subchannel to a particular user. An appropriate labelling of the allocated subchannels is given by a pair $(k,\ell) = \pi(j)$, where k indicates the user the subchannel established in the jth place is assigned to and $0 \leq \ell \leq \min\{t, r_k\}$ identifies that particular subchannel among the subchannels assigned to user k. The function

$$\pi : \{1,\ldots,q\} \to \{1,\ldots,K\} \times \{1,\ldots,t\}$$

assigns to any index j, which indicates encoding order and can be as high as the number of available spatial dimensions given by the rank q of matrix $\boldsymbol{H}$, the subchannel (k,ℓ) allocated at step j. Any subchannel $\pi(j)$ is completely characterized by a transmit weighting vector $\boldsymbol{v}_{\pi(j)} \in \mathbb{C}^{t\times 1}$ and a receive weighting vector $\boldsymbol{u}_{\pi(j)} \in \mathbb{C}^{r_k \times 1}$. The output of the CZF-SESAM algorithm comprises the function π, which indicates the encoding order of each subchannel, and the set of transmit and receive weighting vectors, which characterize the allocated subchannels.

The algorithm works as follows. After having established the first $j-1$ spatial subchannels, the projection matrix $\boldsymbol{T}_j$ is computed as

$$\boldsymbol{T}_j = \boldsymbol{T}_{j-1} - \boldsymbol{v}_{\pi(j-1)}\boldsymbol{v}_{\pi(j-1)}^{\mathrm{H}}$$

with $\boldsymbol{T}_1 = \boldsymbol{I}_t$. This matrix represents the projector of the subspace defined by the intersection of the kernels of the subchannels already established.

Then, channel matrices of all users are projected into this subspace

$$\boldsymbol{H}_k^j = \boldsymbol{H}_k\boldsymbol{T}_j, \quad \forall\, k$$

and SVDs of all projected channel matrices are performed

$$\boldsymbol{H}_k^j = \boldsymbol{U}_k^j\boldsymbol{\Lambda}_k^j\boldsymbol{V}_k^{j,\mathrm{H}}, \quad \forall\, k.$$

Within the subspace defined by $\boldsymbol{T}_j$ the scalar subchannel with maximum gain is characterized by the right and left singular vectors with largest singular value. Following the logic of our ZF-SESAM approach, that subchannel is selected,

$$(k_0, s_0) = \underset{k,s}{\operatorname{argmax}}\left\{\lambda_{k,s}^j\right\}, \quad \pi(j) = (k_0, \ell(k_0)),$$
$$\boldsymbol{v}_{\pi(j)} = \boldsymbol{V}_{k_0}^j\boldsymbol{e}_{s_0}, \quad \boldsymbol{u}_{\pi(j)} = \boldsymbol{U}_{k_0}^j\boldsymbol{e}_{s_0}$$

where $\lambda_{k,s}^j$ is the sth eigenvalue in the main diagonal of matrix $\boldsymbol{\Lambda}_k^j$, $\boldsymbol{e}_{s_0}$ is the s_0th column of $\boldsymbol{I}_{q(k_0,j)}$, being $q(k_0,j)$ the rank of $\boldsymbol{H}_{k_0}^j$, and $\ell(k_0)$ denotes the number of subchannels provisionally assigned to user k_0. At this point the same procedure is repeated to allocate the $(j+1)$th spatial subchannel. An outline of this algorithm is provided in Table III. For subchannel $\pi(j)$ interference caused by subchannels $\pi(i > j)$ is forced to zero since vectors $\boldsymbol{v}_{\pi(i>j)}$ lie within the kernel of this subchannel, i.e.

$$\boldsymbol{u}_{\pi(j)}^{\mathrm{H}}\boldsymbol{H}_{k(j)}\boldsymbol{v}_{\pi(i>j)} = 0,$$

where $k(j)$ indicates the user to which subchannel $\pi(j)$ belongs. By contrast, interference caused by subchannels $\pi(i < j)$ is, in general, not eliminated by the choice of transmit weighting vectors. It will be neutralized by coding. An exception occurs when $k(j') = k(j)$ with

TABLE III
SUCCESSIVE ALLOCATION ALGORITHM FOR CZF-SESAM

```
initialization :  j = 1,  T_1 = I_t
repeat :
  1. H_k^j = H_k T_j ∀k
  2. H_k^j = U_k^j Λ_k^j V_k^{j,H} ∀k
  3. (k_0, s_0) = argmax_{k,s} {λ_{k,s}^j},  π(j) = (k_0, ℓ(k_0))
     v_{π(j)} = V_{k_0}^j e_{s_0},  u_{π(j)} = U_{k_0}^j e_{s_0}
  4. T_{j+1} = T_j − v_{π(j)} v_{π(j)}^H,  j = j + 1
until  j > q
```

$j' \neq j$. In such case it can be shown that subchannels $\pi(j)$ and $\pi(j')$ are entirely decoupled. In fact, the algorithm results in a singular value decomposition if applied to a single user scenario. Effective transmission of information occurs over each of the allocated scalar subchannels whose gain is given by

$$g_{\pi(j)} = \boldsymbol{u}_{\pi(j)}^{\mathrm{H}} \boldsymbol{H}_{k(j)} \boldsymbol{v}_{\pi(j)}.$$

Over this set of virtually decoupled channels sum rate is maximized by allocating the available transmit power according to a waterfilling strategy. As already mentioned, this algorithm, as well as all other discussed in this section, allows a simultaneous and independent execution over each of the subcarriers of an OFDM system, which represents a further practical advantage of these approaches as compared to the optimum solution approaches discussed in Section III-C.

V. PERFORMANCE

Performance of the approaches discussed above has been evaluated and compared by means of simulations.

Fig. 1 shows average sum capacity curves for a Rayleigh distributed channel with $t = 4$ transmit antennas, $K = 2$ users and $r_1 = r_2 = 2$ antennas at each receiver. The entries in the composite channel matrix $\boldsymbol{H}(n)$ corresponding to any subcarrier n have been assumed to be mutually independent with variance equal to one. The horizontal axis represents the ratio between transmit power per subcarrier and the noise variance, which is assumed to be equal for all receive antennas. We have considered a multicarrier transmission system with $N = 16$, in which channels of different subcarriers are assumed mutually uncorrelated. The basic difference between this setting and a system with $N = 1$ is basically the additional degrees of freedom that the multicarrier system offers for allocation of power over the spectral components. Performance of both systems would be identical if a uniform allocation of power were performed over the subcarriers.

Beside the curves corresponding to the approaches discussed above, for comparison purposes, we have also included the following curves. The Sato upper bound on sum capacity, which is reached by the optimum solution [20]. The sum rate achieved by the downlink multiuser MIMO decomposition technique independently reported in [41], [42] and [43], which linearly suppresses interuser interference and is labelled block-ZF. The sum rate achieved by using the column vectors of $\boldsymbol{H}^{-1}$ normalized to norm one as transmit weighting vectors. This approach, which is labelled ZF, linearly suppresses both interuser interference and cross-talk between receive antennas. Finally, two curves corresponding to OFDM access strategies with (OFDMA CSIT) and without (OFDMA no CSIT) channel knowledge at the transmitter.

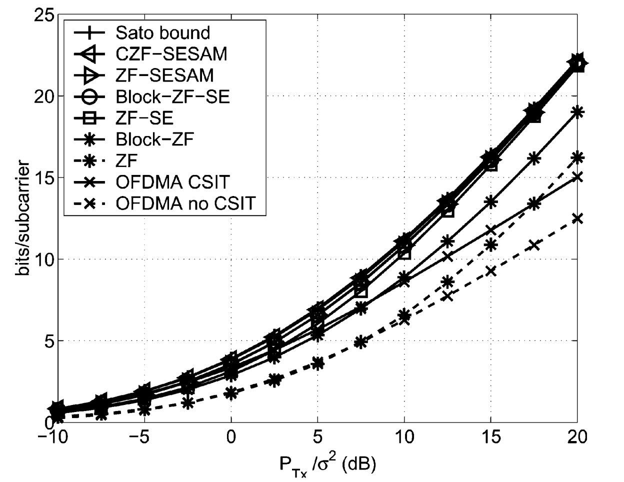

Fig. 1. Average sum capacity for a multiuser setting with spatially uncorrelated Rayleigh fading Gaussian channels. $t = 4$, $r_1 = 2$, $r_2 = 2$, $K = 2$, $N = 16$.

Both techniques serve only one user on every subcarrier. "OFDMA no CSIT" selects the user randomly as no channel knowledge is available at the transmitter. "OFDMA CSIT" selects the user with maximum Frobenius norm. For ZF-SE and block-ZF-SE encoding order has been randomly chosen.

Successive encoding techniques converge to the Sato bound at high values of transmit power. This is in agreement with the result reported in [14], and mentioned before in this section, that states that, if the composite channel matrix $\boldsymbol{H}$ has full row rank, sum capacity achieved by a ZF-SE approach converges to the capacity of the single user MIMO channel represented by the same matrix. Performance of CZF-SESAM overlaps with the Sato bound. Block-ZF-SE and ZF-SESAM show a similar performance with minimal losses with respect to CZF-SESAM due to no optimization of encoding order and no use of cooperation, respectively. As ZF-SE makes use of none of these two degrees of freedom, its performance loss is larger, but still slight.

Linear zero-forcing techniques show a significant performance loss with respect to successive encoding approaches. This is due to the larger number of orthogonality constraints imposed on the choice of transmit weighting vectors as compared to successive encoding techniques, which counteract part of the interference by coding. The performance gap observed between the block-ZF and ZF approaches is due to the larger number of constraints imposed by the latter as a means to suppress not only interuser interference but also cross-talk between receive antennas of a same user. In spite of these constraint-induced losses, at high SNR, the slope of curves corresponding to linear approaches is the same as the slope of curves corresponding to successive encoding approaches. This is not surprising as the asymptotic slope is determined by the number of subchannels over which information is transmitted, which, for both linear and non linear approaches, is equal to the rank of the composite channel matrix $\boldsymbol{H}$.

By contrast, in OFDMA approaches only one user is served on each subcarrier, and therefore, a maximum of two spatial dimensions are used, which represents half of the rank of an average realization of the composite matrix. Accordingly, OFDMA curves asymptotically grow approximately half as fast as all other curves.

In Fig. 2 average sum capacity curves are shown for a scenario as described by the settings used in Fig. 1 but where correlation has been introduced between transmit antenna elements. A transmit correlation matrix $\boldsymbol{R}_{\mathrm{Tx}} = \mathrm{E}\{\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\}$ has been considered with the following eigenvalue profile

$$\boldsymbol{\Lambda} = \mathrm{diag}[14.6254, 1.3525, 0.0220, 0.0001].$$

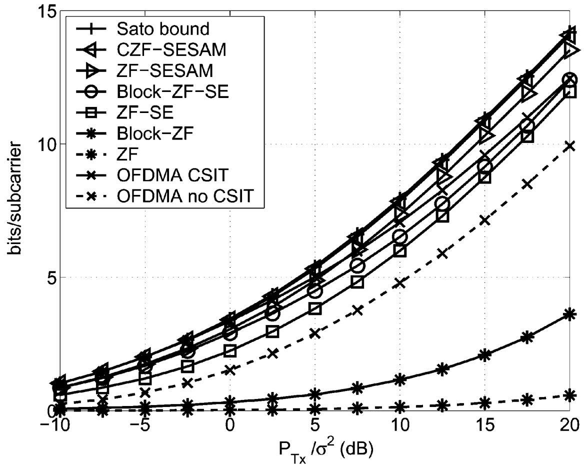

Fig. 2. Average sum capacity for a multiuser setting with spatially correlated Rayleigh fading Gaussian channels. $t = 4, r_1 = 2, r_2 = 2, K = 2, N = 16$.

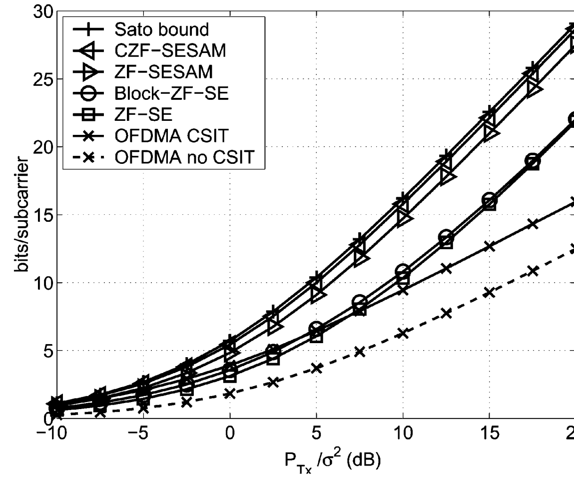

Fig. 3. Average sum capacity for a multiuser setting with spatially uncorrelated Rayleigh fading Gaussian channels. $t = 4, r_k = 2, K = 10, N = 16$.

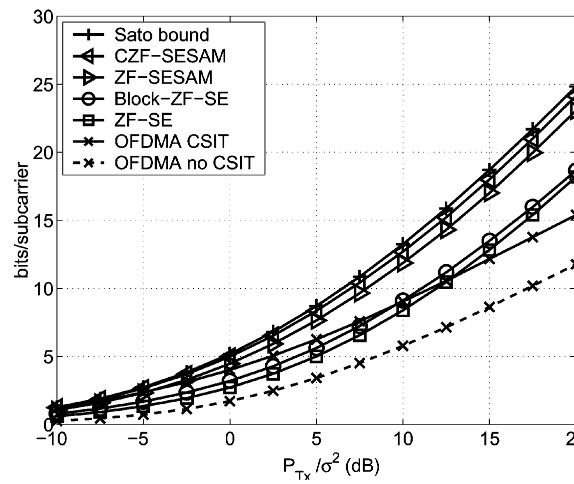

Fig. 4. Average sum capacity for a multiuser setting with spatially correlated Rayleigh fading Gaussian channels. $t = 4, r_k = 2, K = 10, N = 16$.

The practical case of two users being in locations few meters apart from each other that are reached by the base station through quite a narrow bundle of angles of departure matches the setting proposed here.

The asymptotic slope of all curves decay due to the reduced rank of the channel. As particular realizations of composite channel matrices are, at least numerically, not any more full row rank, non-cooperative successive encoding approaches do not approach optimality at high SNR. As before, CZF-SESAM overlaps with the Sato upper bound. The losses of all other successive encoding techniques due to no cooperation and no ordering optimization become larger. Ordering is now crucial since the first selected subchannel largely determines performance as it imposes severe constraints on subsequent subchannels. Cooperation is also important since constructive combination of receive signals raises the gain of the first subchannel. In the light of the simulation results, optimization of encoding order seems to provide more benefit than cooperation between receive antennas. However, the impact of cooperation on capacity will increase with increasing number of receive antennas per user [44]. Linear zero-forcing techniques dramatically suffer from the reduced rank of the channel. Finally, it can be observed that the asymptotic slopes of OFDMA strategies do not strongly differ from the slopes of successive encoding approaches. Indeed, the number of subchannels is, in this case, mostly limited by the rank of the composite channel, which for most realizations is not larger than one or two, and not by the fact of serving only one user on each subcarrier.

Fig. 3 shows average sum capacity curves for a Rayleigh distributed channel with $t = 4$ transmit antennas, $K = 10$ users and $r_k = 2$ antennas at each receiver. Entries in the composite channel matrix of each subcarrier are assumed mutually independent and with covariance equal to one. The number of subcarriers in the system is $N = 16$ and channels corresponding to different subcarriers are assumed mutually uncorrelated. Different from the settings of Figs. 1 and 2, now, the total number of receive antennas in the system is larger than the number of transmit antennas. This calls for a decision regarding the users to be served and the number of subchannels to be assigned to these users on a particular subcarrier. This additional degree of freedom is exploited by the two approaches with successive allocation capability, i.e., CZF-SESAM and ZF-SESAM, and yields a significant performance gain with respect to block-ZF-SE and ZF-SE, which randomly select any two users. CZF-SESAM shows an insignificant loss with respect to the optimum approach. Also moderate is the gain due to receive antenna cooperation of CZF-SESAM with respect to ZF-SESAM. This gain would however increase if receivers with more than two antennas were considered. Linear zero-forcing approaches are not directly applicable to this setting as the number of orthogonality constraints required to linearly suppress interference exceeds the number of dimensions of the transmit weighting vectors. Nevertheless, these techniques might be endowed with a mechanism to preselect a particular group of receive antennas so as to guarantee their applicability. This possibility has not been considered here and, therefore, curves of linear zero-forcing approaches are not included. The slower asymptotic growth of OFDMA strategies with respect to successive encoding strategies can be observed again.

Fig. 4 shows average sum capacity curves for a scenario as described by the settings used in Fig. 3 but where correlation has been introduced between transmit antenna elements. For these simulations, the following eigenvalue profile of the transmit covariance matrix has been considered

$$\mathbf{\Lambda} = \mathrm{diag}[9.6645, 4.9001, 1.2398, 0.1957].$$

This profile may very well match a scenario in which a group of users located in a same certain area, such as a square or street, are reached from the base station over the same moderately broad bundle of angles of departure.

TABLE IV
AVERAGE NUMBER OF ITERATIONS NEEDED BY ITERATIVE ALGORITHM [22] TO ACHIEVE 0.999 $R_{\text{CZF-SESAM}}$

P_{Tx}/σ^2 (dB)

	-10	**-7.5**	**-5**	**- 2.5**	**0**	**2.5**	**5**	**7.5**	**10**	**12.5**	**15**	**17.5**	**20**
Uncorrelated, K = 2	1.1	1.1	1.1	1.2	1.3	1.5	1.6	1.6	1.6	1.5	1.3	1.2	1.1
Uncorrelated, K = 10	1.1	1.1	1.1	1.2	1.7	2.0	2.3	3.1	4.0	4.9	5.5	5.9	6.2
Correlated, K = 2	55.4	37.9	22.7	8.5	4.3	3.2	3.0	3.3	3.9	4.8	5.4	5.6	5.0
Correlated, K = 10	2.6	2.6	2.5	2.6	2.5	2.5	2.6	2.8	2.8	2.9	3.1	3.5	3.8

Again, the moderate rank loss of the channel causes a decay of the asymptotic growth of all approaches. As in Fig. 3, CZF-SESAM shows an insignificant performace loss with respect to the optimum solution and a modest performance gain with respect to ZF-SESAM. Also compared to Fig. 3, the gap between techniques with and without successive allocation capability remains approximately equal while the gap between successive encoding techniques and OFDMA strategies diminishes as the asymptotic growth of the latter is not limited by the rank of the composite channel matrix and, as a result, is practically not affected by the rank reduction due to correlation.

As it can be observed in all four plots, our CZF-SESAM approach practically achieves the performance of the optimum solution while it involves significantly less complexity, as we discussed in Section III. However, this advantage does not come at no cost. Considering a multicarrier transmission system with K users and N subcarriers, we already saw that sum capacity can be achieved in $(K!)^N$ different ways corresponding to all possible combinations of encoding orders on every subcarrier. This represents a degree of freedom that can be used to consider further aspects of system design such as user requirements or fairness. On the contrary, with the CZF-SESAM the ordering is already optimized and we will not have as much flexibility if we want to keep close to the sum capacity of the system. Despite that, simulation results have recently shown that at least for two users the rate region achieved by CZF-SESAM is in most scenarios almost as large as the actual capacity region [45].

In any case, CZF-SESAM as well as all other zero-forcing approaches constitute an efficient method to diagonalize the broadcast channel resulting in a set of virtually decoupled scalar subchannels. These can be used as a platform over which bit and power loading can efficiently be optimized according to the most diverse criteria, possibly at the expense of sum rate. Note that the optimum solution yields a set of mutually coupled channels due to interference. Also the largest singular value criterion used for the successive allocation can be modified if, for instance, an even distribution of spatial dimensions over the users is to be guaranteed. In sum, the CZF-SESAM approach is very versatile and lends itself to be used for system design under a wide range of different criteria, being sum rate just a particular case.

Table IV shows the average number of iterations required by the optimum iterative algorithm of Table I in order to reach 99.9% of the sum rate achieved by CZF-SESAM $(R_{\text{CZF-SESAM}})$. Numbers range between almost one iteration for uncorrelated scenarios at low SNR and more than 50 iterations at -10 dB, $K = 2$ and correlated channels. This indicates that the additional computational complexity of optimal iterative approaches relative to CZF-SESAM strongly depends on the particular setting.

VI. CONCLUSION

In this work allocation methods have been investigated that aim at a maximization of the sum rate transmitted over a broadcast channel. For the sum capacity achieving solution a method has been presented that transforms the resulting vector channels into a set of independent scalar subchannels while preserving capacity, which is meaningful for practical purposes. Also an extension of this solution to a multicarrier system has been proposed. Due to the high complexity of this solution, suboptimum techniques have been investigated that are more easily implementable. To this end, the standard zero-forcing with successive encoding and the SVD approach, which apply to systems with non cooperative and cooperative receive antennas, respectively, have been reviewed. Combining the basic principles of these two techniques and elaborating on the encoding order issue of successive encoding approaches, a novel zero-forcing approach for broadcast channels with groups of cooperating antennas has been presented that is based on successive encoding and a successive allocation of subchannels to users. Simulation results have shown that this technique achieves a larger sum rate than any other state-of-the-art zero-forcing technique closely approaching the Sato upper bound on sum capacity. This bound is reached by the optimum solution with considerably more complexity.

APPENDIX

Proof of Theorem 1: Let be $k_1, k_2 \in \{1,\ldots,K\} \setminus \mathcal{S}_j$ so that $\|\boldsymbol{h}_{k_1}^{\mathrm{T}}\boldsymbol{T}_{j+1}\| \geq \|\boldsymbol{h}_{k_2}^{\mathrm{T}}\boldsymbol{T}_{j+1}\|$. Define $S_{j+1} = S_j \cup \{k_1\}$ and $\mathcal{S}'_{j+1} = \mathcal{S}_j \cup \{k_2\}$ and two ordering functions π and π' such that $\pi(i) = \pi'(i)\ \forall\, i \leq j$, $\pi(j+1) = k_1$ and $\pi'(j+1) = k_2$. We shall then prove $C_{\mathcal{S}_{j+1}} \geq C_{\mathcal{S}'_{j+1}}$.

Assume that the optimum waterfilling allocation for the set $\mathcal{S}'_{j+1}$ yields the waterfilling level $\eta_{\pi'}$ (cf. (20)). For the set $\mathcal{S}_{j+1}$, consider a suboptimum power allocation where the users $\pi(i \leq j)$ are waterfilled to the level $\eta_{\pi'}$ and user k_1 receive the same power as user k_2 in the set $\mathcal{S}'_{j+1}$. Obviously, the first j users achieve the same transmission rate in both sets. Provided that power assigned to users k_2 and k_1 is different from zero, transmission rate achieved by user k_1 will be larger than that achieved by user k_2 as the channel gain of user k_1 is larger than the channel gain of user k_2. If power is zero the maximum transmission rate of both users is zero. Waterfilling allocation of power over the set $\mathcal{S}_{j+1}$ can only lead to an even larger transmission rate $C_{\mathcal{S}_{j+1}}$.

Proof of Theorem 2: First, assume that there is enough transmit power so that all users in $\mathcal{S}_{j+2}$ are allocated some power. For this case, maximizing capacity is equivalent to maximizing (22) and since the order of the first j selected users is fixed the proof reduces to demonstrating

$$\frac{1}{g_{\pi(j+1)}^2} + \frac{1}{g_{\pi(j+2)}^2} \geq \frac{1}{g_{\pi'(j+1)}^2} + \frac{1}{g_{\pi'(j+2)}^2}. \tag{23}$$

By definition we have

$$\begin{aligned} g_{\pi(j+1)}^2 &= \|\boldsymbol{h}_{k_1}^{\mathrm{T}}\boldsymbol{T}_{j+1}\|^2,\\ g_{\pi'(j+1)}^2 &= \|\boldsymbol{h}_{k_2}^{\mathrm{T}}\boldsymbol{T}_{j+1}\|^2,\\ g_{\pi(j+2)}^2 &= \left\|\boldsymbol{h}_{k_2}^{\mathrm{T}}\left(\boldsymbol{T}_{j+1} - \frac{\boldsymbol{T}_{j+1}\boldsymbol{h}_{k_1}^{*}\boldsymbol{h}_{k_1}^{\mathrm{T}}\boldsymbol{T}_{j+1}}{\|\boldsymbol{T}_{j+1}\boldsymbol{h}_{k_1}^{*}\|^2}\right)\right\|^2 \qquad (24)\\ g_{\pi'(j+2)}^2 &= \left\|\boldsymbol{h}_{k_1}^{\mathrm{T}}\left(\boldsymbol{T}_{j+1} - \frac{\boldsymbol{T}_{j+1}\boldsymbol{h}_{k_2}^{*}\boldsymbol{h}_{k_2}^{\mathrm{T}}\boldsymbol{T}_{j+1}}{\|\boldsymbol{T}_{j+1}\boldsymbol{h}_{k_2}^{*}\|^2}\right)\right\|^2 \qquad (25) \end{aligned}$$

and extracting $\|\boldsymbol{h}_{k_2}^{\mathrm{T}}\boldsymbol{T}_{j+1}\|^2$ and $\|\boldsymbol{h}_{k_1}^{\mathrm{T}}\boldsymbol{T}_{j+1}\|^2$ out of the norm operator in (24) and (25), respectively, we obtain

$$g_{\pi(j+2)}^2 = \gamma g_{\pi'(j+1)}^2,$$
$$g_{\pi'(j+2)}^2 = \gamma g_{\pi(j+1)}^2$$

where $\gamma = (1 - \|\boldsymbol{h}_{k_1}^{\mathrm{T}}\boldsymbol{T}_{j+1}\boldsymbol{h}_{k_2}^*\|^2 / \|\boldsymbol{T}_{j+1}\boldsymbol{h}_{k_2}^*\|^2 \|\boldsymbol{T}_{j+1}\boldsymbol{h}_{k_1}^*\|^2)$. Substituting these expressions in (23) and observing $0 < \gamma < 1$ the result follows.

Now, assume that transmit power decreases so that the water-filling level associated with the ordering π sinks below $g_{\pi(j+2)}^{-2} = 1/\gamma g_{\pi'(j+1)}^2$. For the same transmit power, let $\eta_{\pi'}$ be the water-filling level corresponding to the ordering π' and P the amount of optimally allocated power to users k_1 and k_2 for this ordering. Consider a suboptimum power allocation for the channels resulting from ordering π as follows. Users $\pi(i \le j)$ are waterfilled to level $\eta_{\pi'}$ and power P is optimally distributed between users $k_1 = \pi(j+1)$ and $k_2 = \pi(j+2)$ according to a water-filling strategy. With this power allocation, capacity of users $\pi(i \le j)$ is equal for both orderings. In the following we prove that the sum rate achieved by users k_1 and k_2 with the ordering π and under the suboptimum power allocation described above is larger than the sum rate achieved by these same users with ordering π' and the optimum waterfilling power allocation.

If $P > g_{\pi(j+2)}^{-2} - g_{\pi(j+1)}^{-2}$ the proof reduces to demonstrating (23), which has already been done.

If $|g_{\pi'(j+1)}^{-2} - g_{\pi'(j+2)}^{-2}| < P \le g_{\pi(j+2)}^{-2} - g_{\pi(j+1)}^{-2}$, we define $P' = g_{\pi(j+2)}^{-2} - g_{\pi(j+1)}^{-2}$ and the power increment $\Delta P = P' - P$. Let $C_\pi(k_1, k_2)$ be the sum rate achieved by users k_1 and k_2 with ordering π and power P, and $C_{\pi'}(k_1, k_2)$ the sum rate achieved by these same users with ordering π'. According to these definitions

$$C_\pi(k_1,k_2) - C_{\pi'}(k_1,k_2) = \log\left(\frac{(\eta_0 - \Delta P)\eta_0}{\left(\eta_1 - \frac{\Delta P}{2}\right)^2}\right) \tag{26}$$

where $\eta_1 = \left(g_{\pi'(j+2)}^{-2} + g_{\pi'(j+1)}^{-2} + g_{\pi(j+2)}^{-2} - g_{\pi(j+1)}^{-2}\right)/2$ is the waterfilling level for π' when $P = P'$ and $\eta_0 = g_{\pi(j+1)}^{-2} + P' = g_{\pi(j+2)}^{-2}$. The capacity difference in (26) is nonnegative if and only if

$$f(\Delta P) = (\eta_0 - \Delta P)\eta_0 - \left(\eta_1 - \frac{\Delta P}{2}\right)^2 \ge 0. \tag{27}$$

If $g_{\pi'(j+1)}^{-2} \ge g_{\pi'(j+2)}^{-2}$, the maximum power increment is given by $\Delta P_{\max} = g_{\pi(j+2)}^{-2} - g_{\pi(j+1)}^{-2} + g_{\pi'(j+2)}^{-2} - g_{\pi'(j+1)}^{-2}$, and

$$\begin{aligned} f(\Delta P_{\max}) &= \left(g_{\pi'(j+1)}^{-2} - g_{\pi'(j+2)}^{-2}\right)\left(g_{\pi(j+2)}^{-2} - g_{\pi'(j+1)}^{-2}\right) \\ &\ge 0 \end{aligned} \tag{28}$$

if $g_{\pi'(j+2)}^{-2} \ge g_{\pi'(j+1)}^{-2}$, $\Delta P_{\max} = g_{\pi(j+2)}^{-2} - g_{\pi(j+1)}^{-2} - g_{\pi'(j+2)}^{-2} + g_{\pi'(j+1)}^{-2}$, and

$$\begin{aligned} f(\Delta P_{\max}) &= \left(g_{\pi'(j+2)}^{-2} - g_{\pi'(j+1)}^{-2}\right)\left(g_{\pi(j+2)}^{-2} - g_{\pi'(j+2)}^{-2}\right) \\ &\ge 0. \end{aligned} \tag{29}$$

From (28) and (29) and observing that $f(\Delta P)$ is monotonically decreasing the result follows.

Finally, if $P \le \left|g_{\pi'(j+1)}^{-2} - g_{\pi'(j+2)}^{-2}\right|$, all power P is assigned to only one of the two users $\pi'(j+1)$ and $\pi'(j+2)$. Since $g_{\pi'(j+1)}^2 \le g_{\pi(j+1)}^2$ and $g_{\pi'(j+2)}^2 \le g_{\pi(j+1)}^2$, user $\pi(j+1)$ achieves necessarily a higher transmission rate.

The optimum water-filling power allocation over ordering π can only provide larger rates.

References

[1] K. Tachikawa, "A perspective on the evolution of mobile communications," *IEEE Commun. Mag.*, pp. 66–73, Oct. 2003.

[2] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: Design, performance, and field trial results," *IEEE Commun. Mag.*, pp. 143–149, Sep. 2002.

[3] T. A. Thomas, K. L. Baum, and P. Sartori, "Obtaining channel knowledge for closed-loop multi-stream broadband MIMO-OFDM communications using direct channel feedback," *Proc. IEEE GLOBECOM*, 2005.

[4] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 18–28, Jan. 2004.

[5] ——, "User ordering and power allocation for optimal multiantenna precoding/decoding," in *ITG Worksh. Smart Antennas*, Munich, Mar. 2004.

[6] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener filter for the downlink of TDD DS-CDMA systems," in *Proc. ISSSTA 2002*, Prague, Sep. 2002, pp. 9–13.

[7] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," in *ITG SCC'04*, Erlangen, Jan. 2004, pp. 387–394.

[8] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, pp. 2700–2712, Aug. 2005.

[9] M. Bengtsson, "A pragmatic approach to multi-user spatial multiplexing," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, Washington, DC, Aug. 2002.

[10] J. Chang, L. Tassiulas, and F. Rashid-Farrokhi, "Joint transmitter receiver diversity for efficient space division multiaccess," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 16–27, Jan. 2002.

[11] X. Zhang and B. Ottersten, "Joint space-frequency optimization in multiuser MIMO systems," in *IEEE Symp. Signal Process. Inf. Theory (ISSPIT)*, Darmstadt, Germany, 2003.

[12] A. Pascual, A. I. Perez, and M. A. Lagunas, "Joint transceiver optimization in wireless multiuser MIMO-OFDM channels based on simulated annealing," in *Proc. XI European Signal Process. Conf. (EUSIPCO)*, Toulouse, France, 2002.

[13] K. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Adaptive antennas at the mobile and base stations in an OFDM/TDMA system," *IEEE Trans. Commun.*, vol. 49, pp. 195–206, Jan. 2001.

[14] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1706, Jul. 2003.

[15] ——, "On achievable rates in a multi-antenna broadcast downlink," in *Proc. 38th Annu. Allerton Conf. Commun., Contr., Comput.*, Oct. 2000, pp. 1188–1193.

[16] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439–441, May 1983.

[17] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. Inf. Theory*, vol. 24, pp. 374–377, May 1978.

[18] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, pp. 306–311, May 1979.

[19] W. Yu and J. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. IEEE GLOBECOM*, San Antonio, TX, Oct. 2001, pp. 1344–1348.

[20] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

[21] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. Conf. Inf. Sci. Syst.*, 2004.

[22] S. Vishwanath, W. Rhee, N. Jindal, S. Jafar, and A. Goldsmith, "Sum power iterative waterfilling for Gaussian vector broadcast channels," in *Proc. ISIT 2003*, Yokohama, Japan, Jul. 2003, pp. 467–467.

[23] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1570–1580, 2005.

[24] W. Yu, "Sum-capacity computation for the Gaussian vector broadcast channel via dual decomposition," *IEEE Trans. Inf. Theory*, vol. 52, pp. 754–759, 2006.

[25] C. H. F. Fung, W. Yu, and T. J. Lim, "Multi-antenna downlink precoding with individual rate constraints: Power minimization and user ordering," in *Proc. 9th IEEE Singapore Int. Conf. Commun. Syst. (ICCS)*, 2004.

[26] ——, "Precoding for the multi-antenna downlink: Multiuser SNR gap and optimal user ordering," *IEEE Trans. Commun.*, 2006, submitted for publication.
[27] J. Oh, S. J. Kim, R. Narasimhan, and J. M. Cioffi, "Transmit power optimization for Gaussian vector broadcast channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2005.
[28] G. Ginis and J. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 2000, pp. 1627–1631.
[29] C. Windpassinger, T. Vencel, and R. F. H. Fischer, "Precoding and loading for BLAST-like systems," in *Proc. IEEE Int. Conf. Commun.*, Anchorage, AS, 2003, pp. 3061–3065.
[30] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2524–2530, May 1998.
[31] W. Yu, D. P. Varodayan, and J. M. Cioffi, "Trellis and convolutional precoding for transmitter-based interference presubtraction," *IEEE Trans. Commun.*, vol. 53, pp. 1220–1230, 2005.
[32] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1250–1276, Jun. 2002.
[33] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice-strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.
[34] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronic Letters*, vol. 7, pp. 138–139, Mar. 1971.
[35] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 774–780, Aug. 1972.
[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[37] L. Vandenberghe, S. Boyd, and S. P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 19, pp. 499–533, 1998.
[38] P. Tejera, W. Utschick, G. Bauch, and J. A. Nossek, "Efficient implementation of successive encoding schemes for the MIMO OFDM broadcast channel," in *Proc. IEEE Int. Conf. Commun.*, Istanbul, Turkey, 2006.
[39] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–596, Nov. 1999.
[40] H. W. Wolniansky, G. J. Foschini, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Poroc. URSI Int. Symp. Signals, Syst. Electron.*, New York, 1998, pp. 295–300.
[41] R. L. Choi and R. D. Murch, "A downlink decomposition transmit preprocessing techniques for multi-user MIMO systems," in *Poroc. IST Mobile and Wireless Telecommun. Summit 2002*, Thessaloniki, Jun. 2002.
[42] Q. H. Spencer and M. Haardt, "Capacity and downlink transmission algorithms for a multiuser MIMO channel," in *Proc. 36th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2002.
[43] M. Rim, "Multiuser downlink beamforming with multiple transmit and receive antennas," *Electron. Lett.*, vol. 38, pp. 1725–1726, Dec. 2002.
[44] P. Tejera, W. Utschick, G. Bauch, and J. A. Nossek, "A novel decomposition technique for multiuser MIMO," in *Proc. IEEE/ITG Workshop on Smart Antennas*, Duisburg, Apr. 2005.
[45] J. Brehmer, A. Molin, P. Tejera, and W. Utschick, "Low complexity approximation of the MIMO broadcast channel capacity region," in *Proc. IEEE Int. Conf. Commun.*, Istanbul, Turkey, 2006.

The Effect of System Load on the Existence of Bit Errors in CDMA With and Without Parallel Interference Cancelation

Remco van der Hofstad, Matthias Löwe, and Franck Vermet

***Abstract*—In this correpsondence, we study a lightly loaded code-division multiple-access (CDMA) system with and without multistage hard- and soft-decision parallel interference cancelation (HD-PIC and SD-PIC). Throughout this paper we will only consider the situation of a noiseless channel, equal powers and random spreading codes. For the system with no or a fixed number of steps of interference cancelation, we give a lower bound on the maximum number of users such that the probability for the system to have no bit-errors converges to one. Moreover, we investigate when the matched filter system, where parallel interference cancelation is absent, has bit errors with probability converging to one. This implies that the use of HD-PIC and SD-PIC significantly enhances the number of users the system can serve.**

***Index Terms*—Bit-error analysis, code-division multiple access, large deviations, matched filter, number of users, parallel interference cancellation.**

I. INTRODUCTION

The third generation of mobile communication systems has refreshed the interest in so-called code-division multiple access (CDMA), which in third generation (3G) systems is used to increase the communication capacity of the third generation mobile communication systems (see [23] for the history of CDMA). In CDMA systems, all users use the full time window and band of frequencies. Interference with other users is reduced by a coding scheme in which each user multiplies his signal with an individual coding sequence. At the receiver, the signal is retrieved by taking the inner product of the transformed total signal with the corresponding coding sequence. We refer to [27] for general background on CDMA multiuser detection.

While theoretically the coding sequences would best be orthogonal, for practical purposes this is often not necessary and not practical. The orthogonality is thus replaced by "almost orthogonality" as provided e.g., by pseudorandom sequences. This immediately raises the question how many users the system might host, or, in other words, for how many users orthogonal or "almost orthogonal" spreading sequences have the same performance. Of course, the answer to this question is intimately related to the way the interference of signals is canceled. Apparently, the most promising and practical technique are soft- and hard-decision parallel interference cancellation (SD-PIC and HD-PIC), which will be introduced in the next section.

The purpose of the present note is to investigate the CDMA model with spreading sequences of length n and a number of users $k = k_n$ that is of order $O(\frac{n}{\log n})$ or even larger. Here and in what follows we will write k instead of k_n whenever there is no danger of confusion. We

Manuscript received February 17, 2005; revised May 19, 2006. The work of R. van der Hofstad was performed in part at Microsoft Research, and was supported in part by the Netherlands Organisation for Scientific Research (NWO).
R. van der Hofstad is with the Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands (e-mail: rhofstad@win.tue.nl).
M. Löwe is with the Institut für Mathematische Statistik, Fachbereich Mathematik und Informatik, Universität Münster, 48149 Münster, Germany (e-mail: maloewe@math.uni-muenster.de).
F. Vermet is with the Departement de Mathematiques, Universite de Bretagne Occidentale, F-29238 Brest Cedex 3, France (e-mail: Franck.Vermet@univ-brest.fr).
Communicated by X. Wang, Associate Editor for Detection and Estimation.
Digital Object Identifier 10.1109/TIT.2006.881697